Combinatorial Spaces And Order Topologies

Philon Nguyen

ABSTRACT. An archetypal problem discussed in computer science is the problem of searching for a given number in a given set of numbers. Other than sequential search, the classic solution is to sort the list of numbers and then apply binary search. The binary search problem has a complexity of $O(logN)$ for a list of N numbers while the sorting problem cannot be better than $O(N)$ on any sequential computer following the usual assumptions. Whenever the problem of deciding partial order can be done in $O(1)$, a variation of the problem on some bounded list of numbers is to apply binary search without resorting to sort. The overall complexity of the problem is then $O(logR)$ for some radius R. The following upper-bound for finite encodings is shown:

$$O(\log|\mathbf{X}|_\infty)$$

Also, the topology of orderings can provide efficient algorithms for search problems in combinatorial spaces. The main characteristic of those spaces is that they have typical space complexities of $O(2^N)$, $O(N!)$ and $O(N^N)$. The factorial case describes an order topology that can be illustrated using the combinatorial polytope . When a known order topology can be combined to a given formulation of a search problem, the resulting search problem has a polylogarithmic complexity. This logarithmic complexity can then become useful in combinatorial search by providing a logarithmic break-down. These algorithms can be termed as the class of search algorithms that do not require read and are equivalent to the class of logarithmically recursive functions.

1. Computable Structures

A few terminological remarks can be made on the algebra of computable structures.

REMARK 1.1. A computable space is defined using the distinguishability problem.

REMARK 1.2. A computable space is said to be complete if it can represent a Turing-complete space.

Furthermore, the following can be noted:

REMARK 1.3. The equality operator is defined using the cut metric of Equation 3.0.11. This is denoted $=\!\!\!\!\!\underset{\mathcal{N}}{}$.

Whenever $\mathcal{N} = 0$, the conventional equality operator can be used.

REMARK 1.4. Let $\mathbf{T}^n$ be a sequence such that $|\mathbf{T}| = n$. A computable sequence is such that the following holds:

$$(1.0.1) \qquad\qquad f(\mathbf{T}^n) = \mathbf{S}$$

for some computable function f and for a set $\mathbf{S}$ such that:

$$(1.0.2) \qquad\qquad \mathbf{T}_i^n \neq \mathbf{T}_j^n, \quad \mathbf{S}_i \neq \mathbf{S}_j$$

and:

$$(1.0.3) \qquad\qquad \mathbf{T}_i^n = \mathbf{T}_j^n, \quad \mathbf{S}_i = \mathbf{S}_j$$

REMARK 1.5. Let f be a computable function, then f has a representation in a complete computable space.

DEFINITION 1.6. A domain $\mathbb{F}^n$ on a function f is termed a factorial domain if the following holds:

$$(1.0.4) \qquad\qquad f(\mathbf{x}_i) \underset{\mathcal{N}}{\neq} f(\mathbf{x}_j), \quad \mathbf{x}_i \neq \mathbf{x}_i$$

and:

$$(1.0.5) \qquad\qquad f(\mathbf{x}_i) \underset{\mathcal{N}}{=} f(\mathbf{x}_j), \quad \mathbf{x}_i = \mathbf{x}_i$$

2. Order Topologies

An order can be best described as a subset of the integral line. Let $\mathbf{X}$ define a discrete space such that $f : \mathbf{X} \to \mathbf{Y}$ defines an orderable discrete space $\mathbf{Y}$. $\mathrm{Or}(\mathbf{Y})$ then define the order of $\mathbf{Y}$ and a compact order is such that:

$$(2.0.1) \qquad\qquad \frac{\Delta \mathrm{Or}(\mathbf{Y})}{\Delta f(\mathbf{Y})} = 1$$

Whenever $\mathbf{Y}$ is termed orderable, then the following holds:

$$(2.0.2) \qquad\qquad \mathbf{Y}_i - \mathbf{Y}_j = k$$

for some value k. The difference operator $\Delta \mathbf{Y}$ is then defined on a differentiable representation of $\mathbf{X}$. Equation 2.0.1 defines the sequence of integers $(n, n+1, \ldots, n+k)$ or, alternatively, a sequence of ordered incremental values $(a_0, a_0+\delta, \ldots, a_0+n\delta)$ for some unit incremental operator δ. This can be denoted $\mathrm{Or}_1(\mathbf{X})$. Given an arbitrary order $\mathrm{Or}(\mathbf{Y})$, the following holds:

$$(2.0.3) \qquad\qquad \mathrm{Or}_1(\mathbf{Y}) = \pi(\mathrm{Or}(\mathbf{Y}))$$

for some permutation π. Whenever the permutation can be found *seeminglessly*, it can be said that $\mathrm{Or}_1(\mathbf{Y})$ is equivalent to $\mathrm{Or}(\mathbf{Y})$.

EXAMPLE 2.1. The space of rank orders is given by the set of all permutations of the form $\pi(1, 2, \ldots, n)$ [9]. Figure 1 illustrates the Cartesian representation of an order over a prime factorial domain given by:

$$(2.0.4) \qquad\qquad \mathrm{G}_3^3(\pi_i(a_1, a_2, a_3)) = p_1^{a_{i,1}} p_2^{a_{i,2}} p_3^{a_{i,3}}$$

where the G-operator defines a *g*-code of length 3 on a set of cardinality 3. A candidate distance measure in Cartesian space which preserves $\mathrm{Or} \circ \mathrm{G}_3^3(\pi(1, 2, 3))$ is given by:

$$(2.0.5) \qquad\qquad d(\mathbf{x}, \mathbf{y}) = \|\mathbf{wx} - \mathbf{wy}\|$$
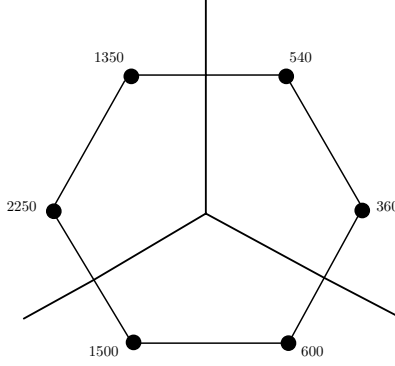
FIGURE 1. Polytope of $\mathrm{Or}(\pi(1,2,3))$ in a prime factorial domain.

for some given weight vector $\mathbf{w}$ and $\mathbf{x}, \mathbf{y} \in \pi(1,2,3)$. It follows that:

$$\mathrm{Or}_1(\pi(1,2,3)) = \pi(\mathrm{Or} \circ \mathrm{G}_3^3(\pi(1,2,3))) \tag{2.0.6}$$

for $\mathbf{p} = (2,3,5)$. Equivalently, an l-code representation can be given as:

$$\begin{aligned} \mathrm{L}_3^3(\pi_i(a_1, a_2, a_3)) = {} & \lfloor a_{i,1} \log p_1 \rfloor_{\log} + \\ & \lfloor a_{i,2} \log p_2 \rfloor_{\log} + \lfloor a_{i,3} \log p_3 \rfloor_{\log} \end{aligned} \tag{2.0.7}$$

where the log-floor operator keeps a logarithmic number of digits in regards to the largest value in the code. In practice, this can be computed using a digit-by-digit radix representation.

EXAMPLE 2.2. Let the following log-prime factorial domain on an index function $\mathcal{I} = (1, \ldots, 2^N)_2$ be defined as:

$$\mathrm{L}_N^n(\mathcal{I}_i) = \sum_{j=1}^N a_i \lfloor \log p_i \rfloor_{\log}, \quad a_i \in (0,1) \tag{2.0.8}$$

It follows that the subsets of the l-codes of Equation 5.0.3 are given by:

$$\begin{aligned} \mathrm{L}_N^n(\mathcal{I}(\mathbf{1}_N \times \mathcal{I}_i)) &= \mathcal{I}_i \mathrm{L}_N^n(\mathcal{I}) \\ &= \sum_{i=j}^N k_j \lfloor \log p_j \rfloor_{\log} \end{aligned} \tag{2.0.9}$$

for $k_j \in (0, \ldots, N)$ and some arbitrary sequence of ordered primes. An illustration of the case $N = 4$ is shown in Figure 2.

Let $\mathbf{X}^N$ define the product space $\mathbf{X} \times \ldots \times \mathbf{X}$ such that $\mathbf{Y}^N \subseteq \mathbf{X}^N$. If there exists some ordering differentiable to one, then then the order is termed cyclical. For some given l-code representation, we have:

$$\frac{\Delta \mathrm{Or}(\pi_{i,j,k\ldots}(\mathrm{L}_N^n(\mathbf{X})))}{\Delta \mathrm{L}(\mathbf{X}}) = 1 \tag{2.0.10}$$

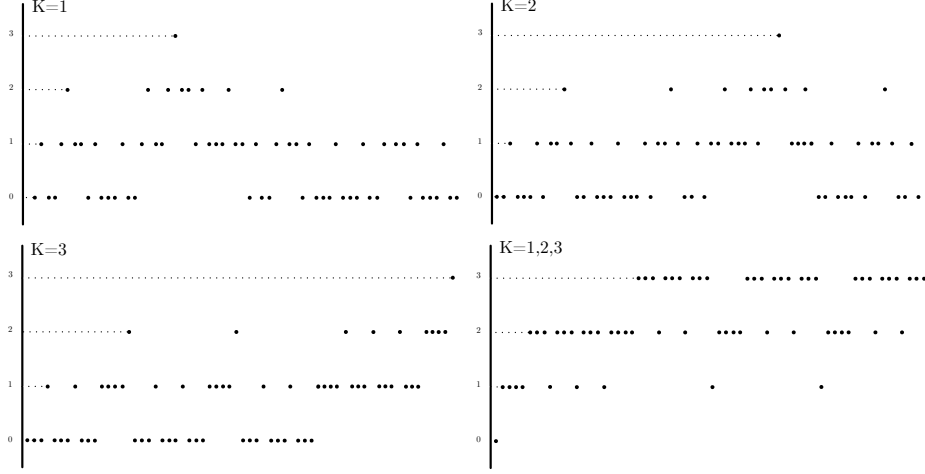given some reshaping permutation operator $\pi_{i,j,k\ldots}$.

FIGURE 2. Order curve.

EXAMPLE 2.3. The indicator function $\mathcal{I} = (1, 2, \ldots, 2^N)_2$ has a (g, l)-code representation such that any point on in $\mathrm{Or}(\mathrm{L}_N^n(\mathcal{I}))$ can be computed using $O(N \log 2)$ recursions, or alternatively, $O(1)$ when the order is known.

EXAMPLE 2.4. The $O(2KM(K + M))$ Jackson-Sheridan-Tseitin transform [**6**, **10**] is given by:

$$(2.0.11) \qquad \mathbf{S} = \left[\begin{array}{c|c} \left(\mathbf{I}_K^T \otimes \mathbf{1}_2\right)^T & \mathbf{B} \\ \hline \mathbf{0} & \left(\mathbf{I}_M^T \otimes \mathbf{1}_2\right)^T \end{array} \right]$$

for some random matrix $\mathbf{B}$ such that $b_{i,j} \in (0, 1)$

3. Distance Equations

Formally, the notions of distance spaces, norms and metrics are defined in the following definitions.

DEFINITION 3.1. A distance measure $d : \mathbf{X} \times \mathbf{X} \to \mathbb{A}$ is a non-invertible mapping that has the following properties:

$$(3.0.1) \qquad\qquad d(\mathbf{x}_i, \mathbf{x}_i) = 0, \ \mathbf{x}_i \in \mathbf{X}$$

$$(3.0.2) \qquad\qquad d(\mathbf{x}_i, \mathbf{x}_j) = d(\mathbf{x}_j, \mathbf{x}_i), \ \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$$

for some non-empty set $\mathbf{X}$ and an alphabet $\mathbb{A}$ isomorphic to an arbitrary subset of $\mathbb{R}_+$, (X,d) defines a distance space.

Given a distance matrix $\mathbf{D} \in \mathbb{R}_+^{N \times N}$ on some finite discrete space $\mathbf{X}$, the mapping $f \circ g$ is invertible whenever $g : \mathbf{X} \times \mathbf{X} \to \Delta$ is invertible. The distance measure on the constructed space is then given by f.

THEOREM 3.2. *A distance space is uniquely defined by* Δ.
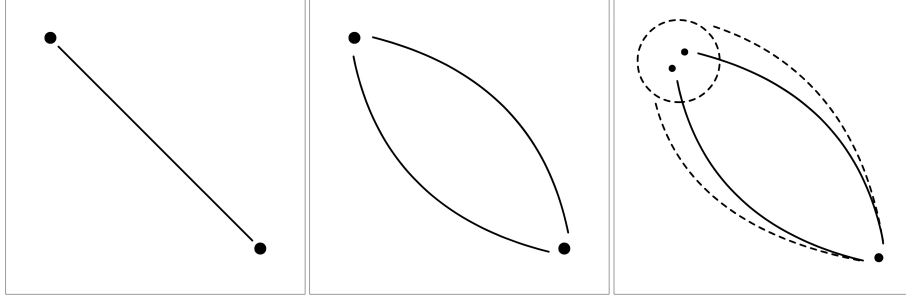
FIGURE 3. Relation diagrams.

By convention, $(\mathbf{X}, d)$ or Δ then denote an arbitrary distance space. It can also be shown that $\Delta \in \mathbb{R}$, $\Delta \in \mathbb{R}^2$, $\Delta \in \mathbb{R}_+^{N \times N+1}$ and $\Delta \in \mathbb{R} \cup \mathbb{R}_+^{N \times N}$ are sufficient, under different assumptions. Any binary function that satisfies Equation 3.0.2 can form a distance space by imposing the following:

$$(3.0.3) \qquad d(\mathbf{x}_i, \mathbf{x}_j) = f'\left(f(\mathbf{x}_i, \mathbf{x}_j), f(\mathbf{x}_j, \mathbf{x}_i)\right)$$

for some function f that satisfies Equation 3.0.2. Other such measures have been used. A distance space embeds a collection of elements onto the real line and induces a poset topology and an order topology.

DEFINITION 3.3. Given some non-empty set $\mathbf{X}$, a metric space $(\mathbf{X}, d)$ is a space equipped with a metric distance measure d defined using Equations 3.0.1 and 3.0.2 with the additional property:

$$(3.0.4) \qquad d(\mathbf{x}_i, \mathbf{x}_k) \leq d(\mathbf{x}_i, \mathbf{x}_j) + d(\mathbf{x}_j, \mathbf{x}_k), \ \ \mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k \in \mathbf{X}$$

Metricity preserves heuristic notions such as distance minimization. Similarly, semimetrics and ultrametrics can be defined by constraining or varying the standard properties.

DEFINITION 3.4. A norm $\|\cdot\|$ is a function $f : \mathbf{X} \to \mathbb{A}$ such that:

$$(3.0.5) \qquad \left\|k\mathbf{x}_i\right\| = \left|k\right|\left\|\mathbf{x}_i\right\|$$

$$(3.0.6) \qquad \left\|\mathbf{x}_i + \mathbf{x}_j\right\| \leq \left\|\mathbf{x}_i\right\| + \left\|\mathbf{x}_j\right\|$$

$$(3.0.7) \qquad \left\|\mathbf{0}\right\| = \mathbb{A}_0$$

for some zero vector $\mathbf{0}$, a zero element $\mathbb{A}_0$, $k \in \mathbf{K}$ and $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$. $\mathbf{X}$ then defines a normed space.

Norms preserve metricity and linearity. Also, the notion of mapping of metric spaces is defined as follows.

DEFINITION 3.5. A mapping $f : \mathbf{X} \to \mathbf{Y}$ of a space $(\mathbf{X}, d)$ into a space $(\mathbf{Y}, d')$ exists if the following holds:

$$(3.0.8) \qquad \frac{1}{k_1} d(\mathbf{x}_i, \mathbf{x}_j) \leq d'(f(\mathbf{x}_i), f(\mathbf{x}_j)) \leq k_2 d(\mathbf{x}_i, \mathbf{x}_j)$$

such that $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$, for some $k_1, k_2 \geq 1$.

An isomorphic mapping is then defined as a (1,1)-distortion mapping. An isometric mapping then defines an isomorphic mapping on a metric space. Mappings of metric spaces can be generalized to mappings of arbitrary spaces using the same definition.

On finite discrete spaces, mappings are given by the set of functions $f : \mathbb{R}_+^{N \times N} \to \mathbb{R}_+^{N \times N}$ such that some parameters (k_1, k_2) hold.

Whenever (k_1, k_2)-distortions can be defined, the distance matrix of the image is said to be bounded by the distance matrix of the domain on some arbitrary function. In the context of similarity search, distortions usually refer to a lower-bounding distortion whenever $k_2 < 1$. An upper-bounding distortion is then given by $k_2 \geq 1$.

In the context of compression, the usual metric is given on some vector space by:

$$(3.0.9) \qquad k = \|f(\mathbf{x}) - \mathbf{x}\|_\infty, \ x \in \mathbf{X}$$

The distortion is then either (k, k), (k_1, k) or (k, k_2).

Typical application domains of mappings include computational geometry, approximation algorithms and functional analysis.

Other problems of practical importance include mappings of random walks into the plane, mappings into binary codes and mappings onto the n-sphere.

A finite discrete space is described in Definition 3.6 and refers to finite computable discrete spaces.

DEFINITION 3.6. Let $f : \mathbf{X} \to \mathbf{Y}$ be a mapping from some arbitrary space $(\mathbf{X}, d)$ into a space $(\mathbf{Y}, d)$. Whenever the following holds for all $\mathbf{y}_i, \mathbf{y}_j \in \mathbf{Y}$:

$$(3.0.10) \qquad \sum_{i,j} d(\mathbf{y}_i, \mathbf{y}_j) = 0$$

given a cut metric defined as:

$$(3.0.11) \qquad d(\mathbf{y}_i, \mathbf{y}_j) = \begin{cases} 1, & \mathcal{N}(\mathbf{y}_i) \cap \mathcal{N}(\mathbf{y}_j) \neq \varnothing \\ 0, & \mathcal{N}(\mathbf{y}_i) \cap \mathcal{N}(\mathbf{y}_j) = \varnothing \end{cases}$$

and some arbitrary neighborhood $\mathcal{N}$, $(\mathbf{Y}, d)$ defines a discrete space. When $\mathbf{Y}$ is finite, f generates a finite discrete space. $(\mathbf{X}, d)$ is then defined as the represented space.

COROLLARY 3.7. *A finite discrete space can be constructed if and only if its elements have an arbitrary neighborhood that can be computed exactly in the classical sense.*

COROLLARY 3.8. *A countable subset of the real line containing arbitrary transcendentals is not a computable discrete space whenever the value of the transcendentals are considered.*

PROOF. Let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ define some space $(\mathbf{X}, d)$ for some transcendentals $\mathbf{x}_i$. Given that the halting problem and the machine equivalence problem are undecidable, the following problem cannot be decided for arbitrary values:

$$(3.0.12) \qquad \mathcal{N}(\mathbf{x}_i) \cap \mathcal{N}(\mathbf{x}_j) = \varnothing$$

$\square$

Equivalently, Equation 3.0.11 cannot be decided on a classical machine for arbitrary subsets of the real line.

COROLLARY 3.9. *The problem of defining any discrete space can be decided upon whenever an arbitrary relation on all its elements can be decided. An arbitrary graph is then defined.*

EXAMPLE 3.10. The space defined by:

$$(3.0.13) \qquad \mathbf{X} = \bigcup_{i=1}^{N} \{\pi + 1/i, 1/i\}$$

is a computable finite discrete space under the following distance matrix:

$$(3.0.14) \qquad \mathbf{D} = \begin{bmatrix} & i=1 & \cdots & 2N & \cdots \\ \hline i=1 & 0 & \cdots & a & \cdots \\ i+1 & b & \cdots & |a-b| & \cdots \\ \vdots & \vdots & \ddots & \vdots & \cdots \\ 2N & a & \cdots & 0 & \cdots \\ 2N+1 & |a+b| & \cdots & b & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

for some element a of an arbitrary alphabet and b given as follows:

$$(3.0.15) \qquad b = \frac{1}{i} - \frac{1}{1+i}$$

such that $a \pm b \in \mathbb{R}_+$.

EXAMPLE 3.11. Let the following arbitrary curve be defined on the following sequence of iterates:

$$(3.0.16) \qquad f_i(x) = f_{i-1}(x) + \delta$$

for some infinitessimal operator δ. Then, f defines a discrete space under the following distance matrix:

$$(3.0.17) \qquad \begin{bmatrix} & f_0(x) & \cdots \\ \hline f_0(x) & 0 & \cdots \\ \vdots & \vdots & \cdots \\ f_i(x) & \sum \delta_i & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

EXAMPLE 3.12. A discrete space can be defined on the following sequence of iterates:

$$(3.0.18) \qquad f_i(x) = f_i(x) + \Omega\delta$$

for some uncomputable number $\Omega \in [0,1]$ [**2**]. A distance matrix can then be given as:

$$(3.0.19) \qquad \begin{bmatrix} & f_0(x) & \cdots \\ \hline f_0(x) & 0 & \cdots \\ \vdots & \vdots & \cdots \\ f_i(x) & \Omega \sum \delta_i & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

4. Distance Equations On Finite Discrete Spaces

Some known notions often associated to search algorithms can be formalized using the distance equation concept.

An example is similarity search in some n-dimensional space equipped with the ℓ_p metric, which has found many interesting algorithmic solutions.

DEFINITION 4.1. Let $\mathbf{X} \cup \mathbf{q}$ be some finite discrete space. A distance equation is defined as follows:

$$(4.0.1) \qquad \delta(\mathbf{q}) : \ d(\mathbf{X}, \mathbf{q}) + k = 0$$

for all values of $k \in \mathbb{K}$ such that $\mathbb{K} \subseteq \mathbb{R}$.

COROLLARY 4.2. *Any distance equation on some finite discrete space defines a set of open or closed intervals.*

COROLLARY 4.3. *A distance equation is defined on a continuous space if and only if $k \in \mathbb{K}$ for some $\mathbb{K} \subseteq \bar{\mathbb{R}}$.*

The complexity of defining a distance equation on a discrete space can then be seen as equivalent to the complexity of defining an equivalent distance equation on a continuous space whenever the corresponding representations of $\mathbf{X}$ are of the same order.

A distance equation is then said to define a k-partition of the real numbers.

DEFINITION 4.4. Two distance equations are isomorphic if they define two k-partitions p, q such that $|p| = |q|$.

DEFINITION 4.5. A composite distance equation is defined using standard set operations on distance equations:

$$(4.0.2) \qquad \left(\bigcup, \bigcap \right) \delta(\mathbf{q}) : \ d(\mathbf{X}, \mathbf{q}) + k = 0$$

The solution of the equation is the subset of X that satisfies the equation and defines subregions of the distance space induced by the order topology.

Whenever the set is viewed as a matrix, it is given by:

$$(4.0.3) \qquad \mathbf{X}_{\mathcal{I}} = \mathcal{I}\mathbf{X}$$

for some indicator function $\mathcal{I}$. And more specifically, by absorption:

$$(4.0.4) \qquad \mathbf{X}_{\mathcal{I}} = \mathcal{I}_{\mathbf{i}}\mathbf{X}_{\mathbf{i}}, \quad \mathbf{i} \subset \mathbb{N}$$

THEOREM 4.6. *Let $\mathbf{X}$ be a finite discrete space, for any distance equation δ defined on a computable distance function, $\mathcal{I}$ is well-defined.*

Also, if δ falls in the complexity class C whenever $\mathcal{I}_i$ can be computed in $O(1)$, then it is customary to denote the resulting complexity class as C^δ.

By extension, any low-complexity computation of $\mathcal{I}_i$ denotes a complexity class C^δ.

Let $\delta_0(\mathbf{q})$ define a distance equation with a solution equal to an arbitrary element chosen uniformly at random from the subset of $\mathbf{X}$ satisfying the adjoint δ equation. This is also referred to as the decision version of the search problem.

The complexity of solving a distance equation δ is equal to a multiple of the complexity of solving its corresponding δ_0 equation. The multiple is in the range $[0, |\mathbf{X}|]$.

4.1. Examples. The similarity search problem in some finite discrete space $\mathbf{X}$ equipped with the ℓ_1 metric is usually defined as follows:

$$(4.1.1) \qquad \delta(\mathbf{q}) : \ \|\mathbf{X} - \mathbf{q}\|_1 \leq \epsilon$$

for some $\epsilon \in \mathbb{R}_+$. The same problem using ℓ_2 metrics is given as:

$$(4.1.2) \qquad \delta(\mathbf{q}) : \ \|\mathbf{X} - \mathbf{q}\|_2 \leq \epsilon$$

Similarly, the orthogonal range search problem can be defined as follows:

$$(4.1.3) \qquad \bigcup_i \delta(\mathbf{q}_i) : \ \|\mathbf{X} - \mathbf{q}_i\|_\infty \leq \epsilon$$

The random convex polytope search problem, also called the point location in arrangement of hyperplanes problem, is done on the discrete space $\mathbf{X}$ defined by the following arrangement:

$$(4.1.4) \qquad \bigcup_i \mathcal{A} : \ \mathbf{A}_i \mathbf{q} \leq \mathbf{b}_i$$

for some non-trivial matrices $\mathbf{A}_i$ with elements $a^i_{j,k}$ and $\mathbf{q}, \mathbf{b} \in \mathbb{R}^n$. Equation 2.1.4 can be translated to a distance equation such as:

$$(4.1.5) \qquad \bigcup_i \delta(\mathbf{q}) : \ \dim \mathbf{q} + \sum_j \mathrm{sgn}^*(\mathbf{A}^i_j q_j - b^i_j) = 0$$

It can also be noted that the sgn* function is here defined on the vertices of the n-cube $\mathbb{H}^n$ as:

$$(4.1.6) \qquad \mathrm{sgn}^* = \left\{ \begin{array}{ll} 1, & x > 0 \\ -1, & x \leq 0 \end{array} \right.$$

The solution of Equation 4.1.5 returns the indexes i of the discrete space $\mathbf{X}$ that solves the distance equation, or alternately the polytopes defined by $(\mathbf{A}_i, \mathbf{b}_i)$.

It has been shown to be equivalent to the search formulation of a Jackson-Sheridan-Tseitin transform.

Let $(\mathbf{Y}, d)$ be some finite discrete space such that $\mathbf{X}_i, \mathbf{Q} \subseteq \mathbf{Y}$. Then, using the Hausdorff distance, the following distance equation can be defined:

$$(4.1.7) \qquad \bigcup_i \delta(\mathbf{Q}) : \ \|\mathbf{X}_i - \mathbf{Q}\|_H + k = 0$$

for $k \in [\epsilon_1, \epsilon_2]$. The Hausdorff distance is defined as usual:

$$(4.1.8) \qquad \|\mathbf{X} - \mathbf{Y}\|_H = \max \left\{ \sup_{\mathbf{x} \in \mathbf{X}} \inf_{\mathbf{y} \in \mathbf{Y}} d(\mathbf{x}, \mathbf{y}), \ \sup_{\mathbf{y} \in \mathbf{Y}} \inf_{\mathbf{x} \in \mathbf{X}} d(\mathbf{x}, \mathbf{y}) \right\}$$

The overall complexity of Equations 4.1.1, 4.1.2 and 4.1.3 can be done, loosely, in at most $O(|\mathbf{X}| \dim \mathbf{X})$, where $|\mathbf{X}|$ denotes the cardinality of the set enclosing the space, following the usual assumptions.

The overall complexity of Equation 4.1.7 is at most $O(|\mathbf{X}||\mathbf{Q}| \dim \mathbf{X})$. On the other hand, Equation 4.1.4 is said to have a search complexity of $O(n^5 \log N)$ [5, 8].

4.2. Classification, Equivalences and Dilations.

4.2.1. *Notation.* For some distance measure d, a search problem denoted δ on some finite discrete space $\mathbf{X}$ that can be solved using d is said to be in the class $\mathcal{C}(\delta, \mathbf{X}, d)$. Also, let the dilation of an arbitrary mapping f of $(\mathbf{X}, d)$ into $(\mathbf{Y}, d')$ be given on the following relation:

$$(4.2.1) \qquad \frac{1}{c_1} \max_i |\mathbf{X}_i| \leq \max_i |\mathbf{Y}_i| \leq c_2 \max_i |\mathbf{X}_i|$$

for some $c_1, c_2 \geq 1$ and for all $\mathbf{q}_i$ and for all subsets $\mathbf{X}_i \subseteq \mathbf{X}$ and $\mathbf{Y}_i \subseteq \mathbf{Y}$ such that:

$$(4.2.2) \qquad \delta(\mathbf{q}_i): \ d(\mathbf{X}_i, \mathbf{q}_i) + k = 0$$

and:

$$(4.2.3) \qquad \delta'(\mathbf{q}_i): \ d'(\mathbf{Y}_i, f(\mathbf{q}_i)) + k' = 0$$

and:

$$(4.2.4) \qquad f(\mathbf{X}_i) \subseteq \mathbf{Y}_i$$

If there is a $(1, O(1))$-dilation, then $\mathcal{C}(\delta, \mathbf{X}, d)$ is said to be equivalent to $\mathcal{C}(\delta', \mathbf{Y}, d')$. The later is denoted as follows:

$$(4.2.5) \qquad \mathcal{C}(\delta, \mathbf{X}, d) \Rightarrow \mathcal{C}(\delta', \mathbf{Y}, d')$$

More specifically, c_2 is expected to grow as follows:

$$(4.2.6) \qquad c_2 = 1 + \frac{\log |\mathbf{X}|}{|\mathbf{X}|}$$

$\mathcal{C}(\delta, \mathbf{X}, d)$ is said to be probabilistically equivalent and bounded to $\mathcal{C}(\delta', \mathbf{Y}, d')$ whenever dilations can be probabilistically bounded. This is denoted:

$$(4.2.7) \qquad \mathcal{C}(\delta, \mathbf{X}, d) \underset{k}{\rightarrow} \mathcal{C}(\delta, \mathbf{Y}, d')$$

Two finite discrete distance spaces $(\mathbf{X}, d)$ and $(\mathbf{Y}, d')$ such that $f : \mathbf{X} \to \mathbf{Y}$ are isomorphically embeddable if the following holds:

$$(4.2.8) \qquad |\mathbf{Y}_i| = |\mathbf{X}_i|$$

for all $\mathbf{q}_i$ and for all subsets $\mathbf{X}_i \subseteq \mathbf{X}$ and $\mathbf{Y}_i \subseteq \mathbf{Y}$ such that Equations 4.2.2, 4.2.3 and 4.2.4 hold. Given two distance measures d and d on some finite discrete spaces $\mathbf{X}$ and $\mathbf{Y}$, if $(\mathbf{X}, d)$ embeds isomorphically into $(\mathbf{Y}, d')$, then $\mathcal{C}(\delta, \mathbf{X}, d)$ is said to be strictly equivalent to $\mathcal{C}(\delta', \mathbf{Y}, d')$. This is denoted as follows:

$$(4.2.9) \qquad \mathcal{C}(\delta, \mathbf{X}, d) \Leftrightarrow \mathcal{C}(\delta', \mathbf{Y}, d')$$

4.2.2. *Examples.* Let the following equation be the general formulation of similarity search on some finite discrete metric space $(\mathbf{X}, d)$ [3–5]:

$$(4.2.10) \qquad \delta(\mathbf{q}): d(\mathbf{X}, \mathbf{q}) \leq \epsilon$$

with $\mathbf{X} \subset \mathbb{R}^n$. Given some $f : \mathbf{X} \to \mathbf{Y}$, it follows that there exists finite discrete spaces $\mathbf{Y}$ for which Equation 4.2.10 is equivalent to is 4.1.1, 4.1.2 and 4.1.3. The set of finite discrete spaces for which this is the case is such that the following holds:

$$(4.2.11) \qquad \begin{aligned} O(\log |\mathbf{X}|) = \max_i |\mathcal{B}^n_\epsilon \left(f(\mathbf{q}_i)\right) \cap \mathbf{Y}| \ - \\ \max_i \left|\mathcal{B}^n_{\epsilon k_1}(\mathbf{q}_i) \cap \mathbf{X}\right| \end{aligned}$$

for all possible values of $\mathbf{q} \in \mathbf{X}$, some ℓ_p ranges $\mathcal{B}^n_\epsilon$ of dimension n and radius ϵ with distortion value k. Also, the following holds whenever a (k_1, k_2)-distortion is defined:

$$(4.2.12) \qquad \begin{aligned} \max_i |\mathcal{B}^n_\epsilon \left(f(\mathbf{q}_i)\right) \cap \mathbf{Y}| \leq \max_i \left|\mathcal{B}^n_{\epsilon k_1}(\mathbf{q}_i) \cap \mathbf{X}\right| - \\ \max_i \left|\mathcal{B}^n_{\epsilon/k_2}(\mathbf{q}_i) \cap \mathbf{X}\right| \end{aligned}$$

Let $f : \mathbb{R}^n_* \to \mathbb{H}^n$ be a mapping to the vertices of the n-cube defining the following nearest neighbor problem:

$$(4.2.13) \qquad f(\mathbf{q}) = \{\mathbf{v} \in \mathbb{H}^n : \ \min_i \|q_i - v_i\|\}$$

The mapping f chooses the vertex of $\mathbb{H}^n$ that is closest in terms of component-wise distances. The point location in arrangement of hyperplanes problem described by Equation 4.1.5 can then be rewritten as follows:

$$(4.2.14) \qquad \delta(\mathbf{q}): \ \dim \mathbf{q} + \dim \mathbf{q} \cos \left(f(\mathbf{q}), \mathbf{1}\right) = 0$$

Therefore, Equation 4.1.5 is strictly yet perhaps trivially equivalent to a linear function of the cosine distance, given some prior nearest-neighbor mapping. Equivalence to ℓ_p can be shown if the following holds:

$$(4.2.15) \qquad \max \|\mathbf{X} - \mathbf{q}\| \leq k$$

for some constant k. The cosine distance is then strictly equivalent to a bounded ℓ_p metric, denoted $\bar{\ell}_p$, which in turn defines a subset of $\mathbb{R}^n$ of finite radius. It is also metrizable by using a convergent sequence that will yield an integral over a domain given by the range of the distance space. The range is then unbounded and cosine distances can be said to be strictly equivalent to ℓ_p metrics for finite discrete spaces. One such mapping for finite vector spaces that can serve this purpose is the limit:

$$(4.2.16) \qquad \cos(\mathbf{x}, \mathbf{y}) = \int_0^{\tan^2 \theta/2} \frac{\|\mathbf{x} - \mathbf{y}\|_p^{1/2}}{1 + \|\mathbf{x} - \mathbf{y}\|_p} \, d\|\mathbf{x} - \mathbf{y}\|_p$$

whenever Equation 4.2.16 generates a computable distance space. As the differential element tends to infinity, the integral yields the range of the cosine. The formulas provided by S. Ramanujan for the computation of π can also serve as an example to the metrization of directional distance spaces.

4.3. Classic Theorems On Discrete Spaces. The class of distance equations that are equivalent or strictly equivalent are tightly coupled with the properties of their adjoint discrete spaces. Known results can be shown in finite vector spaces from the literature on the subject [**5**].

THEOREM 4.7. *In a finite vector space, all norms are equivalent given some mapping $f : \mathbf{X} \to \mathbf{Y}$.*

Theorem 4.7 is denoted:

$$(4.3.1) \qquad \mathcal{C}(\delta, \mathbf{X}, \| \cdot \|_i) \Leftrightarrow \mathcal{C}(\delta, \mathbf{Y}, \| \cdot \|_j)$$

COROLLARY 4.8. *Given any distance measures d and d', there exist finite discrete spaces $\mathbf{X}$ and $\mathbf{Y}$ such that:*

$$(4.3.2) \qquad \mathcal{C}(\delta, \mathbf{X}, d) \Leftrightarrow \mathcal{C}(\delta, \mathbf{Y}, d')$$

PROOF. This can be trivially shown by simple linear manipulations of d and d' on discrete spaces of cardinality two. $\qquad\square$

THEOREM 4.9. [**5**] *Let the cut metric be defined on some subset $\mathbf{X} \subseteq \mathbf{Y}$ as follows:*

$$(4.3.3) \qquad d(\boldsymbol{X}, \boldsymbol{q}) = \left\{ \begin{array}{ll} 1, & f(\boldsymbol{X}, \boldsymbol{q}) = k \\ 0, & otherwise \end{array} \right.$$

Then, some distance metric d is such that:

$$(4.3.4) \qquad \mathcal{C}(\delta, \mathbf{Y}, d) \Leftrightarrow \mathcal{C}(\delta, \mathbf{Y}, \| \cdot \|_1)$$

if and only if d can be expressed as:

$$(4.3.5) \qquad d \equiv \sum_{\mathbf{X} \subseteq \mathbf{Y}} c_i d(\mathbf{X}, \mathbf{q})$$

for some nonnegative and non-zero coefficients c_i.

THEOREM 4.10. *Given an arbitrary distance measure d, a metric d' and some mapping $f : \mathbf{X} \to \mathbf{Y}$, the following holds for some finite vector spaces $\mathbf{X} \subset \mathbb{R}^n$ and $\mathbf{Y} \subset \mathbb{R}^m$ with $m = (n, 3)$:*

$$(4.3.6) \qquad \mathcal{C}(\delta, \mathbf{X}, d) \Leftrightarrow \mathcal{C}(\delta, \mathbf{Y}, d')$$

COROLLARY 4.11. *Given an arbitrary distance measure d and a metric d' some finite vector spaces $\mathbf{X}, \mathbf{Y} \subset \mathbb{R}^n$ and some mapping $f : \mathbf{X} \to \mathbf{Y}$ the following holds:*

$$(4.3.7) \qquad \mathcal{C}(\delta, \mathbf{X}, d) \Leftrightarrow \mathcal{C}(\delta', \mathbf{Y}, d')$$

THEOREM 4.12. *Any finite metric space $(\mathbf{X}, d)$ is probabilistically equivalent and bounded to $(\mathbf{Y}, \| \cdot \|_p)$. This is denoted:*

$$(4.3.8) \qquad \mathcal{C}(\delta, \mathbf{X}, d) \underset{k}{\to} \mathcal{C}(\delta, \mathbf{Y}, \| \cdot \|_p)$$

for some $\mathbf{X} \subset \mathbb{R}^n$ and $\mathbf{Y} \subset \mathbb{R}^{O(\log |\mathbf{X}|)}$.

Theorem 4.12 is such that the relationship between distortions and dilations for the similarity search equation can be given by Equation 4.2.12. Such distortions include low-dimensional subset mappings and random projections.

4.4. Classic Theorems On Subspaces and Subsets. Known theorems on the representation of distance equations over vector spaces as distance equations over subspaces and subsets of vector spaces can be defined from the literature on the subject.

THEOREM 4.13. [**4**] *Let $(\mathbf{X}, d)$ be a finite discrete metric space such that $\mathbf{X} \subset \mathbb{K}^n$ and let $\mathbf{Y}$ be a linear subspace of $\mathbf{X}$ defined as follows:*

$$(4.4.1) \qquad \|\mathbf{y}_i\|_0 = k, \quad \mathbf{y}_i \in \mathbf{Y}$$

and:

$$(4.4.2) \qquad \mathbf{y}_i = \mathcal{I}\mathbf{x}_i, \quad \mathbf{x}_i, \mathbf{y}_i \in \mathbf{X}, \mathbf{Y}$$

for some constant $k \leq n$ and an indicator function $\mathcal{I}$. Then the following holds:

$$(4.4.3) \qquad d(\mathbf{x}_1, \mathbf{x}_2) \geq d(\mathbf{y}_1, \mathbf{y}_2), \quad (\mathbf{x}_1, \mathbf{x}_2), (\mathbf{y}_1, \mathbf{y}_2) \in \mathbf{X}, \mathbf{Y}$$

COROLLARY 4.14. [**7**] *Whenever the indicator functions $\mathcal{I}_i$ describes a contiguous subset of components, the following holds:*

$$(4.4.4) \qquad \delta(\mathbf{q}) \subseteq \bigcup_i \delta'(\mathcal{I}_i\mathbf{q}))$$

given δ and δ' defined as follows:

$$(4.4.5) \qquad \delta(\mathbf{q}): \ \|\mathbf{X} - \mathbf{q}\|_p \leq \epsilon$$

$$(4.4.6) \qquad \bigcup_i \delta'(\mathcal{I}\mathbf{q}_{i,j}): \ \|\mathcal{I}\mathbf{X} - \mathcal{I}_i\mathbf{q}\| \leq O(n^\alpha \epsilon)$$

for $O(n/k)$ indicator functions and a constant $\alpha > 0$.

THEOREM 4.15. [**11**] *Let $\mathbf{X} \cup \mathbf{q} \subset \mathbb{K}^n$ be some finite discrete space, there exist $f, g : \mathbb{R}^n \to \mathbb{R}$ such that the following holds:*

$$(4.4.7) \qquad \bigcup_i \delta(\mathbf{q}_i) \subseteq \bigcup_{i,j} \delta'(f_j(\mathbf{q}_i))$$

given δ and δ' defined as follows:

$$(4.4.8) \qquad \bigcup_i \delta(\mathbf{q}_i): \ \|\mathbf{X} - \mathbf{q}_i\|_\infty \leq \epsilon$$

$$(4.4.9) \qquad \bigcup_{i,j} \delta'(f_j(\mathbf{q}_{i,j})): \ |g(\mathbf{X}) - f_j(\mathbf{q}_{i,j})| = 0$$

THEOREM 4.16. *Let $f : \mathbf{X} \to \mathbf{Y}$ be a distortionless mapping given by a binary matrix $\mathbf{G}$ as follows:*

$$(4.4.10) \qquad f(\mathbf{X}) = \mathbf{G}\mathbf{X}$$

for some $\mathbf{x} \in \mathbf{X}$. Then there exists $\mathbf{G}$ such that following holds for $\mathbf{y}_i, \mathbf{y}_j \in f(\mathbf{X})$ and a constant ϵ:

$$(4.4.11) \qquad \|\mathbf{y}_i - \mathbf{y}_j\|_1 \geq \epsilon, \quad \mathbf{y}_i \neq \mathbf{y}_j$$

$\mathbf{X}$ *and* $\mathbf{Y}$ *can be chosen such that* $\mathbf{X} \subseteq \mathbb{B}_1^n, \mathbf{Y} \subseteq \mathbb{B}_1^m$.

4.5. Characteristic Radius and Encodings. As a converse to the notion of error-correcting bounds, a characteristic radius can be defined on any finite discrete distance space.

DEFINITION 4.17. Given an isomorphic mapping $f : \mathbf{X} \to \mathbf{Y}$, the actual characteristic radius of f is the minimal radius of $\mathbf{Y}$ in the distance space $(\mathbf{Y}, d)$.

This is given by:

$$(4.5.1) \qquad\qquad r = \min d\left(f(\mathbf{x}_i), f(\mathbf{x}_j)\right)$$

for all $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$.

Given a normed finite discrete space $\mathbf{X}$, the compactness c is then defined as:

$$(4.5.2) \qquad\qquad c(\mathbf{X}) = \|\mathbf{X}\|_\infty - r$$

The set of all isomorphic mappings of some finite discrete space is equivalent to the set of all injective mappings $g : f \to \mathbb{R}_+^{N \times N}$ that are order invariant given some distance function $f : \mathbf{X} \times \mathbf{X} \to \mathbb{R}_+^{N \times N}$. For some continuous function $f(x, y, z)$, this can be seen as the smoothing process given by:

$$(4.5.3) \qquad\qquad f(z) = g \circ f(z)$$

for some value z defining an arbitrary cutting plane. Or alternatively:

$$(4.5.4) \qquad\qquad f^{-1} - g \circ f^{-1}\big|_{f(x,y)} = 0$$

which equivalently defines a projection to identity of f and g into $\mathbb{R}^2$.

THEOREM 4.18. *Given some discrete bounded space, a computable characteristic radius can be defined.*

COROLLARY 4.19. *An unbounded discrete space has an actual characteristic radius of ∞.*

COROLLARY 4.20. *Any unbounded discrete space $(\mathbf{X}, d)$ can be bounded through some mapping $f : (\mathbf{X}, d) \to (\mathbf{Y}, d')$.*

Also, most practical results apply to some general finite discrete space or some non-trivial class of such spaces. As the characteristic radius decreases, the complexity of the encoding is said to increase. The complexity is then given by function of the compactness $c(\mathbf{X})$.

For some finite discrete space in $\mathbb{R}$, a few additional results can be obtained.

THEOREM 4.21. *Any finite discrete space $\mathbf{X} \subset \mathbb{R}$ has a characteristic radius upper bounded by:*

$$(4.5.5) \qquad\qquad \log \|\mathbf{X}\|_\infty$$

COROLLARY 4.22. *For some finite value n, any code $\mathbf{X}$ defined in $\mathbb{N}^n$ has a minimal descriptive length upper bounded by $O(\log \|\mathbf{X}\|_\infty)$.*

These results will be further discussed and shown in subsequent sections. Also, Corollary 4.22 can be seen as an information theoretical or algorithmic complexity bound on normed finite discrete spaces. Such bounds are directly proportional to the characteristic radius. A characteristic radius then qualifies how discrete a space is. It defines a compressibility coefficient and by extension a computability order.
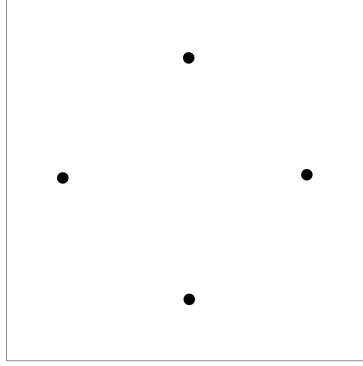
FIGURE 4. Distinguishability on the line.

4.6. Distinguishability On The Line. Let f define some function from some finite discrete space to a finitely computable subset of the real line. The distinguishability problem is then given equivalently to the derivation problem on the real line:

$$(4.6.1) \qquad\qquad f_x = k$$

whenever Equation generates a finite computable set.

It can be shown that the set of functions f that satisfies the distinguishability problem are given by:

$$(4.6.2) \qquad\qquad f_x - f_x^{-1} = k$$

whenever the equation is finite.

Given such a function f, reducing the complexity of the distinguishability problem can be done by finding the set of order preserving mappings of f given by:

$$(4.6.3) \qquad\qquad f^{-1} - g \circ f^{-1}\big|_{f(x)} = 0$$

whenever f is computable.

DEFINITION 4.23. Let f be a norm such that 4.6 holds, then f can be termed a distinguishing norm.

The definition of discreteness in terms of Equation 4.6 can then be restated as follows.

THEOREM 4.24. *Let $f : \mathbf{X} \to \mathbf{Y}$ be an asymptotically finite computable function, then $\mathbf{X}$ is a discrete space.*

The complexity of f defines the computable radius of the discrete space which is conceptually equivalent to the characteristic radius.

5. Polylogarithmic Reduction In Finite Space

Let the g-number G_N^n of a sequence of N integers be defined on some factorial domain.

DEFINITION 5.1. A g-code denotes and by extension defines the set of reversible functions over some factorial domain.

A *g*-code uniquely defines an arbitrary sequence and is equivalent to the set of discrete spaces that can be encoded on a classical computer. Other than its celebrated use in logic and theoretical computer science, the uniqueness property of *g*-codes has been exploited in hashing algorithms. From the perspective of information and quantity, g-numbers typically have large logarithmic norms.

EXAMPLE 5.2. The problem of finding the minimal number of bits required to encode a given sequence of elements from some finite discrete space using *g*-codes is directly related to known information theoretical bounds. Also, polylogarithmic bounds can be specified. This is shown for the particular cases of G_N^1, G_N^8, $\mathrm{G}_N{}^{16,20}$, $\mathrm{G}^{32}{}_N$ and $\mathrm{G}^{64}{}_N$ and their associated distance spaces.

EXAMPLE 5.3. The general problem of determining the compressibility of a sequence has been shown to be uncomputable. More specifically, the algorithmic complexity of a sequence can be reduced to the Halting problem. From an information theoretic point of view, some results show that random sequences cannot improve on the logarithmic scale. Provided specific encodings and distributions, optimal bounds can be effectively be computed. By extension, while exact lower-bounds are in general uncomputable, known upper-bounds have been provided in the literature on the subject.

EXAMPLE 5.4. Polylogarithmic upper-bounds can be provided to the compressibility problem using integral prime *g*-codes. The encoding and decoding of *g*-codes requires the rapid resolution of the constrained prime factorization problem on reasonably integers while the problem of factoring a large integer on a logarithmic scale can be related to the problem of deciding efficiently the integral indeterminate polynomial equation given by:

$$(5.0.1) \qquad\qquad\qquad x - 2^a = 0$$

for some $x \in \mathbb{N}$ and $a \in \mathbb{R}_+$. Whenever a is a rational, the problem can be solved trivially. The decision problem of Equation 5.0.1 is then to determine if an arbitrary power of two accepts an integer solution and corresponds to the integrals solutions of the base two function on a real domain.

Also, it is known that the general solvability of integer-valued indeterminate polynomial equations is an undecidable problem. Also, in its simplest expression given by Equation 5.0.1, the problem is *very* decidable.

An *l*-code is typically defined as follows:

DEFINITION 5.5. The l-number of L_N^n of a sequence of cardinality N is usually defined as:

$$(5.0.2) \qquad\qquad\qquad \mathrm{L}_N^n = f(\mathrm{G}_N^n)$$

for some funtion f.

In general, an *g*-code on some finite discrete space **X** can be defined as:

$$(5.0.3) \qquad\qquad\qquad \mathrm{L}_N^n = f(\mathrm{G}_N^n)$$

such that $r(\mathrm{L}_N^n) = O\left(\log \|\mathrm{G}_N^n\|\right)$ for some given norm.

THEOREM 5.6. L_N^n *defines a finite discrete space equipped with a characteristic radius.*

5.1. L-Codes. From the perspective of compression, common results for g-codes can be discussed for ASCII, UTF, 32-bit encodings and 64-bit encodings.

THEOREM 5.7. *Any prime bitwise g-code has a decoding complexity of $O(N)$ using trial algorithms.*

A bitwise g-code can then be denoted G_N^2. In general, the following complexity result can be stated.

THEOREM 5.8. *Given some basis $f(N) \in \mathbb{K}^N$, a g-code G_N^n has a size complexity upper-bounded by $O(\log \|f(N)\| \|n\|_\infty)$.*

An l-code on ASCII codes can be denoted as L_N^8. The equivalent l-codes for UTF, 32-bit encodings and 64-bit encodings are respectively $L_N^{(16,20)}$, L_N^{32} and L_N^{64}.

EXAMPLE 5.9. Let an l-code L_N^n of length N on an n-dimensional integer code be defined as in Equation 5.0.3 for f defining the logarithmic and inverse functions, a typical ordering of the l-code is then given by, for $N = kn$:

$$(5.1.1) \qquad \begin{aligned} (a_1, \ldots, a_n), ((a_1, \ldots, a_n) + 1) \bmod N, \ldots, \\ ((a_1, \ldots, a_n) + N/n - 1) \bmod N \end{aligned}$$

EXAMPLE 5.10. The ordering of the linear l-codes $L_N^{16} = \sum a_i \log p_i$ and $L_N^{16} = \sum a_i^{-1} \log p_i$ are given by:

$$(5.1.2) \qquad \begin{aligned} (b_1, \ldots, b_{16}) =& (1, 17, 33, 49, 65, 81, 97, 113, 129, 145, \\ & 161, 177, 193, 209, 225, 241) \end{aligned}$$

$$(5.1.3) \qquad \begin{aligned} (b_1, \ldots, b_{16}) =& (241, 225, 209, 193, 177, 161, 145, 129, \\ & 113, 97, 81, 65, 49, 33, 17, 1) \end{aligned}$$

An order curve is then defined as the ordering of some factorial domain function on an indicator function. The order curve of $L_N^n(\mathcal{I})$ is shown in Figure 5.

Example 5.10 defines an oriented lattice. Similar orders can be defined in $\|L_N^n - 0\|_\infty$. The lattices of Example 5.10 have orientation 1 and -1 respectively.

5.2. Jackson-Sheridan-Tseitin Transforms. Let $\mathbf{S}$ be a finite discrete space defining a Jackson-Sheridan-Tseitin transform [6, 10] such that $s_1 < \ldots < s_n$ and:

$$(5.2.1) \qquad \mathbf{S}' = \mathcal{I}(\mathbf{S} \otimes \mathbf{1}_{2^n})$$

It can be shown that the order set of $L_N^n(\mathcal{I})$ generates a quasi-lattice for $N = 2^n$. Let the quasi-lattice $L_N^n(\mathcal{I})$ be denoted $f(x)$ and let the following order curve be defined as $\Delta[f]_i(x)$ for some constant $h = k$. Then the following can be shown:

$$(5.2.2) \qquad \Delta[f]_i(x) = 2^i + 2^{i+1}(i - 1)$$

Or alternatively:

$$(5.2.3) \qquad i \equiv k \bmod 2^k$$

Let S_N^n denote the order curve on $L_N^n(\mathcal{I})$ such that:

$$(5.2.4) \qquad S_N^n = L_N^n(\mathbf{X}) = L_N^n(\mathcal{I}(\mathbf{A} \otimes \mathbf{1}_{2^N}))$$

S_N^n can be defined as follows:

$$
\begin{aligned}
S_N^n &= \sum_{i=1}^{N} \lfloor a_i f(s_i) \log p_i \rfloor_{\log} \\
&= \sum_{i=1}^{N} \lfloor a_i \log p_i \rfloor_{\log} + \Delta S_N^n(\mathcal{I})
\end{aligned}
$$

(5.2.5)

for some function $f(s_i)$, $a_i \in \{0,1\}$, $s_i \in \mathbf{S}$ and a logarithmic code floor function. p_i then defines an arbitrary set of ordered primes.

REMARK 5.11. A set of good constants are such that $f(s_i) \approx 1$. In general, a set of *good* constants are such that $f(s_{i,j,\ldots}) \approx k_{i,j,\ldots}$ such that $|\mathbf{k}|$ is bound (Equation 5.2.10).

Following this remark, whenever $f(s_i) \approx 1$, the order of ΔS_N^n is equal to the order of L_N^n for some l-function such that:

$$
(5.2.6) \qquad\qquad r\left(\Delta S_N^n\right) = o\left(\|L_N^n(\Delta S_N^n)\|_\infty\right)
$$

This is denoted:

$$
(5.2.7) \qquad\qquad \mathrm{Or}(\Delta S_N^n) = \mathrm{Or}(L_N^n)
$$

THEOREM 5.12. *An ordered bounded space of cardinality 2^n can be searched in $O(n \log 2)$ whenever partial order can be decided in $O(1)$.*

THEOREM 5.13. *Let $Or(L_N^n) = Or(S_N^n)$ on some function f defined as in Equation 5.2.5, and let the l-code S_N^n be defined such that the following holds:*

$$
(5.2.8) \qquad\qquad Or(L_N^n(\pi(0,1,2,3,\ldots 2^n - 1)_2)) = Or(S_N^n)
$$

A set of constants $f(c_i s_i) = c_i f(s_i)$ can be found with space complexity $c_i = O(\log \|\mathbf{S}\|_\infty + n \log k)$.

COROLLARY 5.14. *Let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be some finite discrete space with order $Or(\mathbf{X})$. An order preserving function is such that for $\mathbf{F}_{i,j} = f(\mathbf{x}_i)$, there exists $f : \mathbf{F} \to \mathbf{K}^N$ such that $Or(f(\mathbf{X})) = Or(L_N^n(\mathcal{I}))$.*

REMARK 5.15. Let $(\mathbf{S}/T, 1)$ be an arbitrary Jackson-Sheridan-Tseitin transform with solutions $\mathcal{I}_s$. A general form for distance equations that retrieves a set $(L_N^n)_s$ is:

$$
(5.2.9) \qquad\qquad \bigcup_{i=1}^{n} \delta(q = A_i) \colon d(q, B_{i,j}) = 1
$$

such that $j \neq i$ and $j \in \mathcal{I}_s$. And:

$$
(5.2.10) \qquad \bigcup_{i=1}^{n} \delta(q = A_i) \colon d(q, f(B_{i,j}, C_{i,j,k}, \ldots, D_{i,j,\ldots})) = 1
$$

such that $\ldots \neq j \neq i$ and $j, \ldots \in \mathcal{I}_s$. Whenever a set $\mathbf{c}_{i,j}$ of constants can be found such that Remark 5.11 or Theorem 5.13 hold, the solution is said to be efficient.
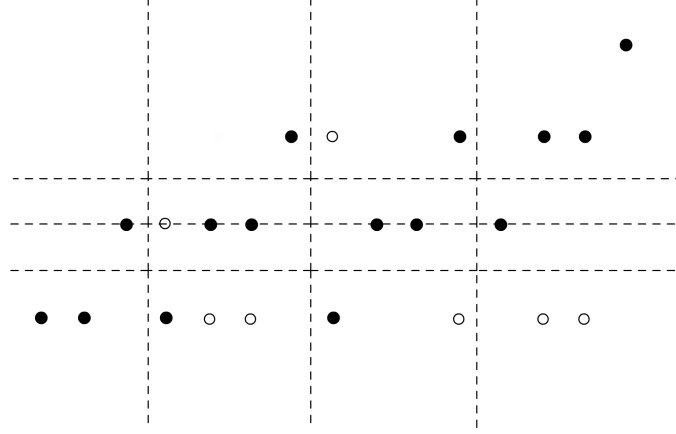
FIGURE 5. Order curve of $L_N^n(\mathcal{I})$.

EXAMPLE 5.16. Starting from the $\|\mathbf{S}\|_0 + \|\mathbf{S}\|_1 = O(2KM\,(K+M))$ Jackson-Sheridan-Tseitin representation, the problem can then be stated as such [**6**, **10**]:

$$(5.2.11) \qquad \mathbf{S}' = \mathcal{I}\left(\mathrm{L}_{K+M}^2(\mathbf{S}) \otimes \mathbf{1}_{2^{2KM}-1}\right)$$

such that $s_{i,j} \in \{0,1\}$ and $\mathbf{s}_i \in \mathbf{S}$. Three l-codes can be constructed. The first one, on the decimal binary representation of $\mathbf{S}$:

$$(5.2.12) \qquad \mathrm{L}_{K+M}^2(\mathbf{S}) = \sum_{i=1}^{K+M} \lfloor \mathcal{I}_i \log p_i \rfloor_{\log}$$

A second one on $f(\mathbf{S}')$:

$$(5.2.13) \qquad \mathrm{L}_{2KM}^2 \circ f \circ \mathrm{L}_{K+M}^2(\mathbf{S}_i) = \sum_{i=j}^{2KM} \lfloor \mathcal{I}_i c_i \mathrm{L}_{K+M}^2(\mathbf{S}_i) \log p_i \rfloor_{\log}$$

such that:

$$(5.2.14) \qquad c_i \mathrm{L}_{K+M}^2(\mathbf{S}_i)\ (\approx, \geq)\ 1$$

This satisfies Remark 5.11 and a candidate set of constants c_i can then be given as:

$$(5.2.15) \qquad \left\lfloor \left(\mathrm{L}_{K+M}^2(\mathbf{S}_i)\right)^{-1}\left(1 + \frac{i}{O((K+M)^{\lfloor \log 10 \rfloor})}\right)\right\rfloor_{\log}$$

A third l-code on $\mathbf{S}'$ is given by:

$$(5.2.16) \qquad \mathrm{L}_{2KM(K+M)}^{2KM+1}(\mathbf{S}') = \sum_{i=1}^{K+M} \lfloor k_i \log p_i,\ \ k_i \in [0,k] \rfloor_{\log}$$

In practice, rational-valued polylogarithmic functions can be used.

The following properties hold on the order curves:

$$(5.2.17) \qquad \mathrm{Or}\left(\mathrm{L}_{K+M}^2(\mathbf{S})\right) \subset \mathrm{Or}\left(\mathrm{L}_N^n(\mathcal{I})\right)$$

$$(5.2.18) \qquad\qquad \text{Or}\left(\text{L}^2_{2KM}(f(\mathbf{S}'))\right) = \text{Or}\left(\text{L}^2_{K+M}(\mathbf{S})\right)$$

$$(5.2.19) \qquad\qquad \text{Or}\left(\text{L}^{2KM+1}_{2KM(K+M)}\right) = \mathbf{A}/\mathbf{B}$$

such that $\mathbf{A} = \{0,\ldots,\prod_{i=1}^{2KM} p_i^k\}$ and $\mathbf{B}$ are the numbers factoring to at least a p_j such that $j \neq i$. The following cardinalities can then be computed:

$$(5.2.20) \qquad \begin{aligned} |\mathbf{A}| &= (K+M+1)^{K+M} \\ |\mathbf{B}| &= n - (K+M+1)^{K+M} \end{aligned}$$

The distribution of $\text{Or}\left(\text{L}^{2KM+1}_{2KM(K+M)}\right)$ behaves as follows:

$$(5.2.21) \qquad \lim_{K+M\to\infty} \frac{\Delta\text{Or}\left(\text{L}^{2KM+1}_{2KM(K+M)}(\mathbf{S}')\right)}{\Delta\text{L}^{2KM+1}_{2KM(K+M)}(\mathbf{S}')} = 1$$

Given Equation 5.2.21, as the problem grows larger, the sparsity of the order curve decreases. Furthermore, the distributions of finite differences has a factorial growth complexity, It can then be said that the space complexity of the associated l-code is given by:

$$(5.2.22) \qquad \left|\text{L}^{2KM+1}_{2KM(K+M)}(\mathbf{S}'_i)\right| = O(k\log\|\mathbf{p}\|_\infty)$$

Following this, a general form of an associated distance equation is given by:

$$(5.2.23) \qquad \bigcup_{i=1}^{2KM} \delta(\mathbf{q} = A_i): \ \mathbf{q} - B_{i,j} = T$$

An actual equation can be given by:

$$(5.2.24) \qquad \begin{aligned} &\bigcup_{i\in\mathcal{I}_s} \delta(q = \sum_{i=1}^{K+M} \lfloor a_i \log p_i \rfloor_{\log}): \ q + \sum_{\substack{j\neq i \\ j\in\mathcal{I}_s}} \sum_{k=1}^{K+M} \left\lfloor a_{j,k} \log_{p_{j,k}} \right\rfloor_{\log} \\ &\approx \sum_{i=1}^{K} \lfloor \log p_i \rfloor_{\log} + 3\sum_{i=K+1}^{M} \lfloor \log p_i \rfloor_{\log} \end{aligned}$$

Alternatively, in l-code notation:

$$(5.2.25) \qquad \bigcup_{i\in\mathcal{I}_s} \delta(q): \ q + \mathcal{I}\text{L}^{2KM+1}_{2KM(K+M)}(\mathcal{I}) = T$$

Since the zero distance never occurs, Equation 5.2.23 can be termed a distance equation. The approximation factor can be set to equality using Theorem 5.22.

PROPERTY 5.17. *A totient function can be given to count the elements in a prime or log-prime order:*

$$(5.2.26) \qquad \phi(n, \boldsymbol{p}) = n \prod_{p_i \in \boldsymbol{p}} \left(1 - \frac{1}{p_i}\right)$$

such that $gcd(n, p_i) \neq 1$.

In practice, arithmetic progressions or some arbitrary ordered sequence of primes are sufficient, or alternatively, a set of p-adic numbers.

This result can be generalized to factorial domains as defined in Definition 1.6.

PROPERTY 5.18. *The following on ordered primes holds:*

$$(5.2.27) \qquad \sum_{i \in \mathcal{I}} \log p_i < \sum_{j \in \mathcal{J}} \log p_j, \quad \mathcal{I} \subset \mathcal{J}$$

REMARK 5.19. Let the following sequence of iterates be defined on the order given by Equation 5.2.19:

$$(5.2.28) \qquad f_i = x, \quad x \in \mathrm{Or}\left(\mathrm{L}_{2KM(K+M)}^{2KM+1}\right)$$

It can be shown that $\log_i \log_i f_i$ is an inverse sigmoidal function.

Following Remark 5.19, a random sample at the extremities of the order can be used to approximate the value $\log_i \log_i f_i$. From the approximated value, the lowest-nearest value to the sequence of iterates can be chosen.

THEOREM 5.20. $Or(L_N^n(\mathcal{I}))$ *can be computed using* $O(\log N)$ *recursions.*

THEOREM 5.21. *On a logarithmic factorial prime domain, the space complexity of a finite discrete space $\boldsymbol{X}$ is upper bounded by:*

$$(5.2.29) \qquad O(\log \|\boldsymbol{X}\|_\infty)$$

THEOREM 5.22. *A code of cardinality k in some distance space is reversible if it provides a k-partition.*

The order on the l-codes of some indicator function can be computed without the knowledge of the associated l-codes. The indicator function defines a space of complexity $O(2^n)$ or $O(n^n)$. The complete search can then be done in approximately $O(n \log 2)$ and $O(n \log n)$ respectively. A relaxation of the problem only requires the knowledge of the number of elements before and after a given element in the l-code. This can be done in $O(1)$. The upper-bound provided on the finite difference of the class of integral prime l-codes then provides a bound on the space complexity of the coding.

5.3. Further Remarks on g-Codes and l-Codes.
5.3.1. *Iterated Encodings.*

DEFINITION 5.23. Let n be the length of a code, N, the number of codes in a (g,l)-code, i, the number of iterations and j the number of (g,l)-codes in a given iteration, an iterated encoding is then recursively defined as follows:

$$(5.3.1) \qquad {}_j^i\mathrm{G}_N^n = {}_{j_1}^{i_1}\mathrm{G}_{N_1}^{n_1} \circ \bigcup {}_{j_2}^{i_2}\mathrm{G}_{N_2}^{n_2} \circ \ldots \circ \bigcup {}_{j_k}^{i_k}\mathrm{G}_{N_k}^{n_k}$$

And similarly:

$$(5.3.2) \qquad {}_j^i\mathrm{L}_N^n = {}_{j_1}^{i_1}\mathrm{L}_{N_1}^{n_1} \circ \bigcup {}_{j_2}^{i_2}\mathrm{L}_{N_2}^{n_2} \circ \ldots \circ \bigcup {}_{j_k}^{i_k}\mathrm{L}_{N_k}^{n_k}$$

Iterated encodings preserve the complexity of a single iteration whenever the number of iterations is $O(1)$. This is denoted:

$$(5.3.3) \qquad ECDC\left({}_j^i\mathrm{G}_N^n\right) = O\left(ECDC\left(\mathrm{G}_{O(\|\mathbf{N}\|_\infty)}^{O(\|\mathbf{n}\|_\infty)}\right)\right)$$

FIGURE 6. Matrix values of $f(\pi(1, 2, 3, 4, 5))$.

5.3.2. *Limiting Encodings.* An infinite sequence is incompressible using g-codes, which follows the classic argument in algorithmic complexity that shows that the problem of defining such a code is in general uncomputable. The logarithmic upper-bound of l-codes follows the classic information theoretical argument.

The basic results of algorithmic complexity can be restated as such for codes and sequences.

THEOREM 5.24. *An infinite code may or may not have a finite algorithmic encoding.*

THEOREM 5.25. *The number of infinite codes that have a finite algorithmic encoding is finite.*

THEOREM 5.26. *The problem of determining the existence of a finite algorithmic encoding for an infinite code is decidable.*

In the context of finite discrete spaces, the main result is as follows.

THEOREM 5.27. *Finite codes have an algorithmic encoding upper-bounded by $\log N$ where N is given by the ℓ_∞ norm.*

5.3.3. *l-Functions.* An l-function is defined as a mapping from an arbitrary encoding to a set of l-codes. Formally:

DEFINITION 5.28. Let l be an l-function. The following then holds:

$$(5.3.4) \qquad l(\mathbf{X}) = \left\{ {}^{i_1}_{j_1}\mathrm{L}^{n_1}_{N_1}, \ldots, {}^{i_k}_{j_k}\mathrm{L}^{n_k}_{N_k} \right\}$$

Reversibility and uniqueness properties can be discussed in the context of l-functions.

DEFINITION 5.29. An l-function is said to be reversible if and only if the following holds:

$$(5.3.5) \qquad f(\mathbf{X}) = \bigcup_k l_k(\mathbf{X}_k)$$

for some bijective function f defining a unique l-code.

5.3.4. *Complete l-Functions.* Complete encodings define as usual the set of encodings that are Turing complete.

THEOREM 5.30. *The following sequence of iterates illustrates a Turing machine:*

$$(5.3.6) \qquad \mathbf{x}_i = \mathbf{x}_{i-1} + \mathbf{r}_i, i = 1, \ldots$$

with the following:

$$(5.3.7) \qquad \bigcup_k l_k(\mathbf{x}_k) = \bigcup_k {}^{i_k}_{j_k} L^{n_k}_{N_k}(\mathbf{x})$$

The l-code representation of a universal encoding is denominated a complete l-function. Theorem 5.30 provides a class of simple l-functions built on the permutation space of an algorithmic encoding.

5.3.5. *Application of l-Function on (1,0).* Let some function f define a sorting algorithm known as bead sort [1] on N random integers $\mathbf{B}$ in their matrix rank order representation such that:

$$(5.3.8) \qquad \bigcup_{k,j=1,N}^{M,1} f_i(b_{j,k}, b_{j+1,k}) = \begin{cases} \begin{matrix} \ddots & \vdots & \cdot^{\cdot^{\cdot}} \\ \cdots & b_{j+1,k} & \cdots \\ \cdots & b_{j,k} & \cdots \\ \cdot^{\cdot^{\cdot}} & \vdots & \ddots \end{matrix}, & (b_{j,k}, b_{j+1,k}) = (1,0) \\ \begin{matrix} \ddots & \vdots & \cdot^{\cdot^{\cdot}} \\ \cdots & b_{j,k} & \cdots \\ \cdots & b_{j,k+1} & \cdots \\ \cdot^{\cdot^{\cdot}} & \vdots & \ddots \end{matrix}, & otherwise \end{cases}$$

and:

$$(5.3.9) \qquad \mathbf{B} = \begin{bmatrix} b_{1,1} & \dots & b_{1,M} \\ b_{1,2} & \dots & b_{2,M} \\ \vdots & \ddots & \vdots \\ b_{N,1} & \dots & b_{N,M} \end{bmatrix}$$

with $b_{i,j} \in \{0,1\}$.

Then the following representation can be used:

$$(5.3.10) \qquad f_i = \sum_{i=1}^{N!} \frac{a_i{}^i}{b_i}$$

Equation 5.3.10 can be denoted as the $\mathcal{L}$-interpolation of f and defines the upper-bound of on the following complexity space:

$$(5.3.11) \qquad \begin{array}{c|c} O(f) & O(|f|) \\ \hline O(f) & O(N!) \\ \vdots & \vdots \\ O(N) & O(1) \end{array}$$

This is shown in Figure 6.

References

[1] J.J. Arunlanandham, C.S. Calude, and M.J. Dinneen, *Bead-sort: a natural sorting algorithm*, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS **76** (2002), 153–162.

[2] Cristian S. Calude and Michael J. Dinneen, *Exact approximations of omega numbers*, I. J. Bifurcation and Chaos **17** (2007), no. 6, 1937–1954.

[3] P. Ciaccia, M. Patella, and P. Zezula, *M-tree: An efficient access method for similarity search in metric spaces*, Proceedings of the 23rd VLDB International Conference (1997), 426–435.

[4] C. Faloutsos, *Searching multimedia databases by content*, 1996.

[5] J.E. Goodman and J. O'Rourke, *Handbook of discrete computational geometry*, 2004.

[6] P. Jackson and D. Sheridan, *Clause form conversions for boolean circuits*, Theory and applications of satisfiability testing, 2005, pp. 183–198.

[7] W.-K. Loh, Y.-S. Moon, and K.-Y. Whang, *Efficient time-series subsequence matching using duality in constructing windows*, AITrc Technical Report (2000).

[8] G. Salton, A. Wong, and C. S. Yang, *A vector space model for automatic indexing*, Communications of the ACM (1975), 613–620.

[9] R.S. Schulman, *A geometric model of rank correlation*, The American Statistician **33** (1979), no. 2, 77–80.

[10] G.S. Tseitin, *On the complexity of derivation in propositional calculus*, Structures in constructive mathematics and mathematical logic, 1968, pp. 115–125.

[11] C. Yu, *High-dimensional indexing: Transformational approaches to high-dimensional range and similarity searches*, Springer, 2002.

6. Acknowledgements

GNetCorp Inc. - Technical Reports